Acta Genet Med Gemellol 35: 35-48 (1986)

Received 15 October 1985
Final 13 January 1986

The Offspring of Twins as Sampling Units in Pedigree Analysis of Congenital Anomalies

Andrew A. Kramer[1], Linda Corey[2]

[1] *Department of Social and Preventive Medicine, State University of New York at Buffalo;* [2] *Department of Human Genetics, Medical College of Virginia, Richmond*

Abstract. A statistical model was developed to determine the likelihood of a twin kinship, that is, the offspring of a pair of monozygotic or dizygotic twins under three types of inheritance: sporadic, single locus fetal genetic, and single locus maternal genetic. Samples of 8,000 kinships were simulated for a discrete trait under various hypotheses, and the likelihood determined for each type of etiology. The results indicated that the pedigree analysis procedures formulated here could efficiently detect sporadic or single locus effects with a power approaching 100%, although the parameter estimates obtained were slightly biased. Further analyses revealed that the type of pedigree analysis formulated in this study was found to have equivalent power for equal or unequal frequencies of kinships by the sex and zygosity of the twin parent. It was suggested that further studies be carried out that included the twins and spouses in the likelihood equations, as well as tests of more sophisticated models.

Key words: Offspring of twins, Pedigree analysis, Congenital anomalies

INTRODUCTION

Various methods have been developed to assess the relative importance of genetic and environmental factors in the determination of observed variation in man, particularly with respect to the presence or absence of disease state or abnormality. Pedigree analysis is one of the more widely used techniques in human quantitative genetics to have emerged in the past decade. This method, however, is limited by the fact that its ability to provide specific parameter estimates is dependent upon the structure of the

particular pedigrees examined. That is, certain parameters are confounded for specific pedigree types and hence are not estimable.

Maternal genetic influences have long been recognized as potentially important sources of variation in multifactorial traits in man. The need for a methodology which will permit the clear distinction between fetal and maternal genetic effects has stimulated the development of techniques using less conventional family structures. A particularly powerful design of this type employs the use of the families of adult monozygotic (MZ) twins to broadly detect maternal influences [11].

Although the MZ twin kinship design was initially developed to examine variation in metric characters, it has been extended to include discontinuous traits such as birth defects [12,13]. In order to distinguish among fetal and maternal genetic effects, however, it may be beneficial to include both the families of dizygotic (DZ) as well as MZ twins into the analysis. This paper attempts to test such a design using birth defect information from twin families in pedigree analysis.

MATERIALS AND METHODS

Formulation of the Models

Three specific hypotheses involving sporadic, fetal genetic, and maternal genetic causation, were considered in this study. Parameter estimates were obtained using a risk function for a model in the form of the logistic distribution, which were derived under each hypothesis. The sporadic model contains only one parameter, the population incidence, while the fetal and maternal models specify four parameters; the gene frequency and three genotypic penetrance functions. The logistic distribution was chosen because it is widely used in epidemiological studies of Bernoulli response variables, such as disease status, and has many appealing factors, such as its sigmoid shape, range of 0 to 1, and the flexibility to add multiple variables that can be of nominal, ordinal, or continuous form [7].

Under a model which assumes a sporadic etiology (Fig. 1) for the trait, every individual chosen at random from the population has the same risk of being affected, regardless of genotype. In this case, individual risk can be written as $(1+e^{-y})^{-1}$, where y is a constant for a given population, such that individual risk is equal to the population incidence. Figure 2 illustrates the situation when the trait is controlled by a single fetal genetic locus with two alleles. The mean values of the three possible genotypes are represented by X_{AA}, X_{Aa} and X_{aa}. Any variance around these mean values would arise from nonfamilial sources. In this case, risk for a particular disorder is represented by the solid curve, and the genotype dependent probability of being affected would be:

$$\Pr(\text{individual affected/genotype i}) = \frac{1}{1+e^{-X_i}}$$

where X_i = genotype specific liability value.

The population incidence then becomes:

https://doi.org/10.1017/S0001566000006255 Published online by Cambridge University Press

$$\text{Incidence} = \sum_{i=1}^{3} p_i \frac{1}{1+e^{-X_i}}$$

where p_i is the genotypic frequency.

The final hypothesis which was examined specifies that the trait in question is under the influence of the mother's genotype, ie, a direct maternal effect (also Fig. 2). This

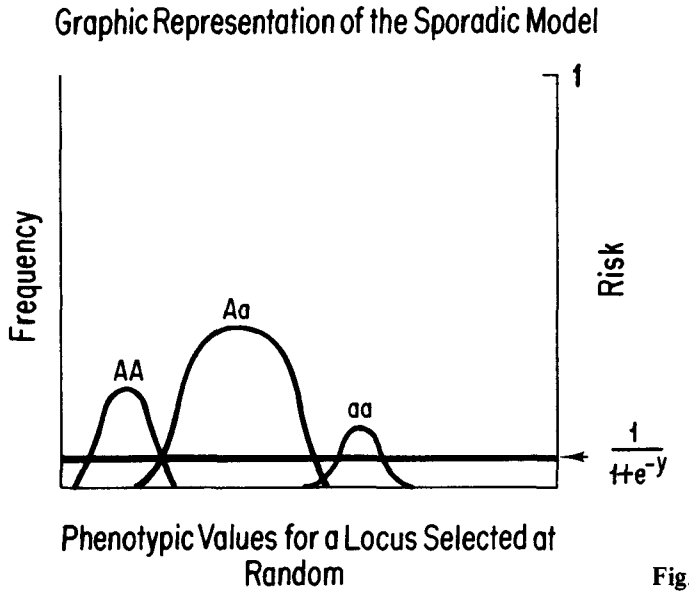

Fig. 1

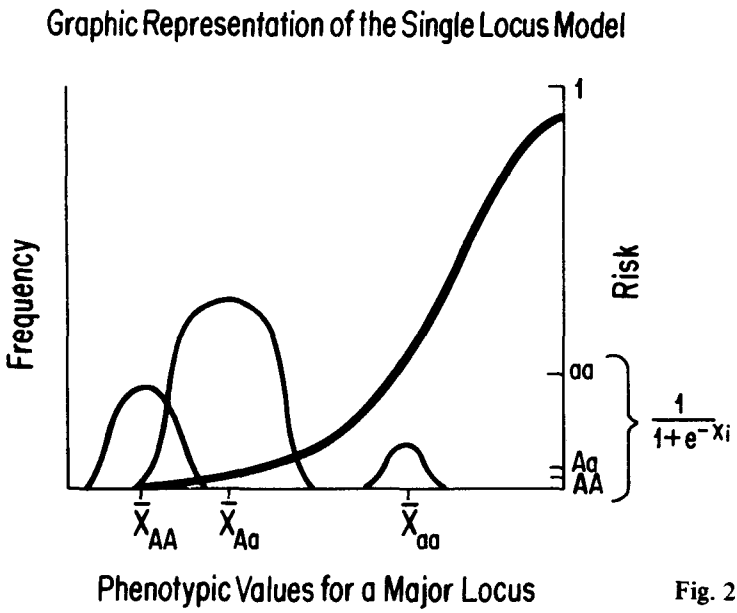

Fig. 2

https://doi.org/10.1017/S0001566000006255 Published online by Cambridge University Press

situation is analogous to the preceding hypothesis, except that the genotypic values observed arise from the maternal genotype, not that of the fetus. The corresponding equations which result are:

$$\Pr(\text{affected/mother has genotype i}) = \frac{1}{1+e^{-M_j}}$$

M_j = maternal genotype specific liability value

$$\text{Incidence} = \sum_{j=1}^{3} s_j \frac{1}{1+e^{-M_j}}$$

where s_j is the maternal genotypic frequency.

Although maternal environmental effects are a potential confounder, they would mimic a genetic effect if they were consistent across a female's pregnancies. If this consistency was specific to one twin, then the fetal genetic model would be accepted. Only in the situation where a maternal environmental effect is concordant for the pregnancies of female twins, but with a higher concordance for MZ females, would it be mistaken for an inherited maternal effect.

Derivation of the Likelihoods

The sampling units used in this study were twin kinships, a kinship being comprised of the offspring of a twin pair. The likelihood of a kinship was obtained from calculating the probability of observing R affected individuals out of S offspring of each twin. In order to arrive at maximum likelihood estimates (MLE) of the parameters, the log likelihood for each family was calculated, then summed over all kinships, and an iterative search routine used to maximize the parameter values with respect to the total likelihood.

Sporadic Etiology

Let X denote the probability that an individual acquires the trait, which is equal to the population incidence. For a kinship with S_1 and S_2 children of whom R_1 and R_2 are affected offspring of twin 1 and twin 2, respectively, the likelihood of a kinship is:

Lik(kinship/sporadic etiology) =

$$\binom{S_1+S_2}{R_1+R_2} X^{(R_1+R_2)} (1-X)^{(S_1+S_2-R_1-R_2)}$$

Fetal Genetic Etiology

Separate likelihoods must be calculated for MZ and DZ kinships, since MZ twins are genetically identical, and DZ twins have on the average only one-half of their genes in common. The model assumes that there is a single gene influencing the trait with two alleles and three genotypes in Hardy-Weinberg equilibrium for both the parental and offspring generation. Some essential notation follows:

ψ_{ij1} = Probability of both spouses and a twin chosen at random from twin pair having genotypes i,j,l

$S_{k/j}$ = Probability of a DZ twin being genotype k given cotwin is genotype j

https://doi.org/10.1017/S0001566000006255 Published online by Cambridge University Press

$P_{c/jk}$ = Probability of a child being genotype c given parents with genotypes j,k

G_c = Genotype specific risk for trait = $\dfrac{1}{1+e^{-X_i}}$

S_1, S_2, R_1, R_2, X_i as above

As with the likelihood for the sporadic model, the genetic likelihood is a multinomial, except that it must be weighted by the parental genotypic probabilities. For MZ twins and their spouses, this is simply the product of the three genotypic probabilities ψ_{ijl} · DZ twins present a more difficult case. In this instance, the genotypic probabilities for the two spouses and one twin must be multiplied by the conditional probability of the cotwin being genotype k given his twin is genotype j. These conditional probabilities were derived by the following equation:

$$S_{k/j} = P(kUj)/P(j) = \sum_{a=1}^{3} \sum_{b=1}^{3} [W_a W_b P_{j/ab} P_{k/ab}] / P(j)$$

where a = father's genotype, b = mother's genotype;
W_i = probability of genotype i

The likelihoods for the offspring of MZ and DZ twins, respectively, are given below:

$$\text{Lik(MZ offspring)} = \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{l=1}^{3} \psi_{ijl} \binom{S_1}{R_1} \binom{S_2}{R_2}$$

$$\left[\sum_{c=1}^{3} P_{c/ij} G_c\right]^{R_1} \left[\sum_{c=1}^{3} P_{c/ij} (1-G_c)\right]^{(S_1-R_1)} \left[\sum_{d=1}^{3} P_{d/jl} G_d\right]^{R_2}$$

$$\left[\sum_{d=1}^{3} P_{d/jl}(1-G_d)\right]^{(S_2-R_2)}$$

$$\text{Lik(DZ offspring)} = \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{l=1}^{3} \psi_{ijl} \sum_{k=1}^{3} S_{k/j}$$

$$\binom{S_1}{R_1} \binom{S_2}{R_2} \left[\sum_{c=1}^{3} P_{c/ij} G_c\right]^{R_1} \left[\sum_{c=1}^{3} P_{c/ij}(1-G_c)\right]^{(S_1-R_1)}$$

$$\left[\sum_{d=1}^{3} P_{d/kl} G_d\right]^{R_2} \left[\sum_{d=1}^{3} P_{d/kl} (1-G_d)\right]^{(S_2-R_2)}$$

Maternal Genetic Etiology

Under the maternal genetic model, separate likelihoods must be calculated for MZ female, DZ female, and male kinships. The likelihoods which are obtained are simpler than for the genetic model, since affected status depends solely on the maternal genotype and no transmission probabilities are needed. As an example, the likelihood for DZ female kinships is listed in the following.

https://doi.org/10.1017/S0001566000006255 Published online by Cambridge University Press

$$\text{Lik(DZ female kinship)} = \sum_{m=1}^{3} \psi_m \sum_{n=1}^{3} S_{n/m} \binom{S_1}{R_1} \binom{S_2}{R_2}$$

$$G_m{}^{R_1}(1-G_m)^{(S_1-R_1)}\, G_n{}^{R_2}(1-G_n)^{(S_2-R_2)}$$

ψ_m = probability of being genotype m

G_m = probability of affected given maternal genotype m $= \dfrac{1}{1+e^{-Mj}}$

$S_{n/m}$ = probability of a DZ twin having genotype n given co-twin is genotype m

S_1, S_2, R_1, R_2, M_j as above

In the actual minimizations for each procedure the log of likelihood is used with the constant $\binom{S_1}{R_1}\binom{S_2}{R_2}$ dropping out of each equation. The log likelihood ratio statistic is used to test various hypotheses, where in comparing two models;

−2 log likelihood (Model A − Model B) is distributed as a chi-square statistics with m degrees of freedom, where m is the difference in the number of parameters under the two hypotheses.

Simulation Procedures

Computer simulations were carried out to test the various statistical properties of the procedure developed in this study. To that end, samples of 8,000 kinships were generated under a given hypothesis with specific parameter values. A kinship was comprised of three offspring for each twin. The sample sizes used for this simulation study were chosen so as to reflect a realistic number of twin pairs which could be obtained from existing large scale twin registries. An iterative search routine from the Numerical Algorithm Group Fortran Library minimized the −log likelihood (−LL) for a parameter set. The −LLs from competing models were compared, and this was repeated 50 times for a balanced design and 20 times for an unbalanced design (unequal number of sex by zygosity proportions of the twin parents).

I. Balanced Design

a. Type I Error. There are two unique problems associated with the hypotheses considered in this study. One problem, which will be explored in detail later, is that both the fetal and maternal genetic models contain the same number of parameters, so no formal statistical test is possible. The second problem lies in the comparison of genetic models against the null hypothesis of a sporadic model. Since the genetic models examined in this study contained three more parameters than the sporadic model, there should be three degrees of freedom in a likelihood ratio test involving these hypotheses. However, this does not hold in the present case, because the alternative hypotheses reduce to the null hypothesis when either the gene frequency is zero or the genotypic effects are equal. Thus, the true number of degrees of freedom are unknown. To that end, samples of 8,000 kinships comprised of 2,000 for each sex by zygosity

category of the twin parents were generated 50 times under the null (sporadic) hypothesis and tested under all hypotheses, with the log likelihood ratio (LR) being calculated for each instance. The population incidence used in these simulations was 0.006, since this corresponded with the frequency of some of the more prevalent birth defects thought to be influenced by maternal effects, such as congenital dislocation of the hip and pyloric stenosis [1,9]. After 50 simulations, the mean of the LRs under each alternative hypothesis was calculated. This figure was then compared with the median point in the chi-square distribution under various degrees of freedom to arrive at the appropriate testing criteria. In addition, the number of significant tests was compared with that expected under various degrees of freedom.

b. Power Studies. Simulations of kinships under the genetic models were done in order to see if the correct model would be selected, and to determine the rejection region for comparing hypotheses with the same degrees of freedom. Again, a balanced design with 8,000 kinships was used. In this simulation, two sets of parameter values were chosen (Table 1) such that the risk for siblings and offspring of affected individuals was

Table 1. Parameter Values Used in the Genetic Simulations of Kinship Data*

	Gene Frequency	Penetrance function (probability)					
		aa		Aa		AA	
Set A	0.045	2.197	(0.90)	−3.127	(0.042)	−7.382	(0.000622)
Set B	0.015	2.197	(0.90)	−1.640	(0.1624)	−6.876	(0.00103)

* For the penetrance function, these values represent the parameters in the logistic risk function.

close to that obtained for birth defects, 3-9%, with a population incidence of 0.006. One hundred simulations were done for each hypothesis, 50 for each of the 2 sets of parameter values.

II. Unbalanced Design

To further determine the efficacy of using twin kinships, kinships were simulated for an unequal number of sex X zygosity twin parents. Estimates of the proportion of kinships for each sex and zygosity category were based on the frequency of various twin types found in a large population-based twin registry. These proportions were multiplied by 8,000, the expected number of complete twin kinships in the registry, to arrive at the final estimates of specific sample sizes. Twenty simulations were carried out with these proportions to provide estimates ot Type I error and power, as described above.

RESULTS

Table 2 shows one of the data sets produced from a simulation under a fetal genetic hypothesis, with a balanced design (2,000 kinships for each parental twin type). The number of kinships indicate how many twin pairs had that specific number of affected

https://doi.org/10.1017/S0001566000006255 Published online by Cambridge University Press

Table 2. Example of a Data Set Simulated Under a Fetal Genetic Hypothesis*

Twin type	Affected children Twin 1 (N)	Affected children Twin 2 (N)	Kinships (N)
MZ-M	0	0	1946
MZ-M	0	1	24
MZ-M	1	0	24
MZ-M	1	1	3
MZ-M	1	2	1
MZ-M	2	0	3
MZ-F	0	0	1931
MZ-F	0	1	27
MZ-F	0	2	1
MZ-F	1	0	37
MZ-F	1	1	2
MZ-F	1	2	1
MZ-F	2	0	1
DZ-M	0	0	1931
DZ-M	0	1	31
DZ-M	0	2	3
DZ-M	1	0	26
DZ-M	1	1	4
DZ-M	2	0	5
DZ-F	0	0	1936
DZ-F	0	1	25
DZ-F	0	2	5
DZ-F	0	3	1
DZ-F	1	0	26
DZ-F	1	1	5
DZ-F	2	0	2

*Three children per twin.

offspring (kinship type). For example, there were 2 MZ female kinships in which each twin had one affected offspring. Although separated in this table, there was no analytical difference between kinship types that were symmetric (eg, twin 1 one affected, twin 2 none and twin 1 none, twin 2 one affected). Since the log of the likelihoods were used, the number of kinships were multiplied by the likelihood for that kinship type and summed across kinship types for a specific hypothesis.

Table 3 gives the -LLs for the three hypotheses tested. Both genetic hypotheses were highly significant when tested against the sporadic model; however, the χ^2 for improved fit was higher by 15.4 for the fetal genetic model.

When distributions of kinships were simulated 50 times under a model of sporadic causation, the majority of LR values were zero or very small, indicating that the incorrect alternative hypotheses could not explain the data any better than did the sporadic hypotesis. To decide the level at which the test of improved fit for an alternative hypothesis should be considered significant, the LR values were compared with the median of the χ^2 distribution and the expected number of significant results corresponding to distributions with one and two degrees of freedom, as shown in Table 4.

Table 3. −Log Likelihoods Under Each Hypothesis for a Specific Fetal Genetic Simulation
(See Table 2)

Hypothesis	−Log likelihood	χ^2 vs. sporadic
Sporadic	1791.1	
Maternal genetic	1744.1	94.0
Fetal genetic	1736.4	109.4

Table 4. Comparison of Results of Sporadic Simulations with Alternative χ^2 Distributions (N = 50)

		Significance at			
Alternative hypothesis	Average log likelihood ratio value	1 df α=0.10	1 df α=0.05	2df α=0.10	2 df α=0.05
Fetal genetic	0.572	5	2	2	0
Maternal genetic	0.654	3	2	2	1

Median $\chi^2_1 = 0.455, \chi^2_2 = 1.386$.

The mean LRs for the fetal and maternal genetic hypothesis were 0.572 and 0.654, respectively. These values are closer to a median χ^2 value of 0.455 for one degree of freedom than 1.386 or 2.366, the median values for two and three degrees of freedom, respectively. Also, the number of the LRs that reached significance for α= 0.10 and α= 0.05 approximated the expected number under one rather than two or three degrees of freedom.

Kinships were generated under a fetal genetic hypothesis for two different sets of parameter values. The null hypothesis of sporadic causation was resoundingly rejected in favor of a fetal genetic model ($P < 0.001$) in every instance. This resulted in a power of 100%, regardless of the parameter values used.

The number of degrees of freedom in the fetal/maternal genetic LR is zero as both procedures estimate the same number of parameters; hence, significance tests cannot be carried out. However, this ratio can be compared against various values to determine the appropriate LR level needed to reach a desired power. The results of such a comparison are presented in Table 5. In order to achieve a power of at least 85-90%, it was necessary to reject the null hypothesis of maternal genetic transmission in favor of the true hypothesis of a fetal genetic effect whenever the LR was greater than zero.

The results of testing a maternal genetic versus either a sporadic or fetal genetic hypotesis on data simulated under a maternal genetic transmission model were excellent. Values for the maternal genetic/sporadic LR were greater than 200 for most trials, and the minimum LR was above 100. The maternal/fetal genetic LR, for which

Table 5. Proportion of Fetal/Maternal Genetic Likelihood Ratios Exceeding Specific Values for Data Simulated Under a Fetal Genetic Model (N = 50)

	> 0.0	3.84*	7.88 **
Set A	0.88	0.66	0.52
Set B	0.94	0.80	0.64

* Critical value for $\alpha = 0.05$, 1 df.
** Critical value for $\alpha = 0.005$, 1 df.

there is no distribution-based statistical test, was generally quite high. Further, a rejection level of 0.0 would result in a power of 100% with α=0.05; that is, the correct hypothesis of a maternal causation would have always been selected.

The values of the final parameter estimates obtained from minimizing the likelihood under a fetal genetic model for a single simulation depended heavily on the initial values used, with the result that the parameter estimates differed from their true values in absolute (but not relative) magnitude. Minimization under the maternal genetic model always yielded a single, and presumably global minimum for different initial parameter values, although again the parameter estimates were slightly deviant.

Distributions of kinships were simulated 20 times under a sporadic hypothesis for unequal sex by zygosity proportions of twins (Table 6). This was done to test twin kin-

Table 6. Sex by Zygosity Classification of Parents Used in Unbalanced Design

	Sex		
Zygosity	Male	Female	Total
MZ	0.196	0.234	0.430
DZ	0.263	0.307	0.570
Total	0.459	0.541	1.000

ships as sampling units in pedigree analysis in a more realistic situation, that is, when there are unequal proportions of twin parents by sex and zygosity categories. The resultant likelihood ratio rests of a genetic versus a sporadic hypothesis are shown in Table 7. The fetal and maternal genetic hypotheses were each accepted on two occasions when the chi square statistic was assumed to have one degree of freedom. When data simulated under a fetal genetic model with unequal number of sex by zygosity kinships were evaluated, the sporadic model was rejected in every case. As shown in Table 8, the proportion of fetal genetic/maternal genetic LRs exceeding certain values correspond fairly closely to those using a balanced design (Table 5). Tests involving unbalanced data simulated under the maternal genetic hypothesis produced LRs greater than 200 when tested against the sporadic model. Further, the maternal/fetal genetic LRs were greater than 10 for each run, and generally above 70 and 25 for parameter sets A and B, respectively.

Table 7. –2 Log Likelihood Ratio Values for Unbalanced Data Simulated Under a Sporadic Model

Run	Fetal hypothesis	Maternal hypothesis	Run	Fetal hypothesis	Maternal hypothesis
1	0	0	11	0	0
2	0	0	12	0	0
3	0	0	13	0	0
4	0	0	14	0	0
5	0.60	0.20	15	0.60	0.20
6	6.20**	3.20	16	0	0
7	0	0.20	17	0	0
8	0	0	18	4.00*	4.20*
9	0	0.20	19	0	1.00
10	3.20	5.00*	20	0	0

Mean value: Fetal hypothesis = 0.73, Maternal hypothesis = 0.76.
Number of kinships: 1566 MZM, 1869 MZF, 2102 DZM, 2463 DZF.
* Significant at $\alpha = 0.05$, 1 df.
** Significant at $\alpha = 0.05$, 2 df.

Table 8. Proportion of Fetal/Maternal Genetic Likelihood Ratios for Unbalanced Data Simulated Under a Fetal Genetic Model Exceeding Specific Values (N = 20)

	> 0.0	3.84*	7.88**
Set A	0.95	0.70	0.50
Set B	0.95	0.90	0.80

* Critical value for $\alpha = 0.05$, 1 df.
** Critical value for $\alpha = 0.005$, 1 df.

DISCUSSION

Simulation studies were carried out on kinships comprised of the offspring of twins randomly generated under a sporadic hypothesis of no genetic effects in order to determine the appropriate number of degrees of freedom for tests involving that hypothesis. The results obtained suggested that the –2 log likelihood ratio of a fetal or maternal genetic hypothesis versus a sporadic hypothesis is distributed as a χ^2 with one degree of freedom. Thus, in tests of hypotheses using the pedigree analysis procedures developed here, the correct number of degrees of freedom which should be used is one, rather than simply the difference in the number of parameters estimated.

Both of the fetal and maternal genetic models which were examined had a power of 100% when tested against the sporadic hypothesis. Although 8,000 twin kinships were simulated, the χ^2 values were of such magnitude that a much smaller number of kinships may be acceptable. It is also quite possible that the use of twin kinships might

allow the detection of fetal or maternal genetic effects for multifactorial traits with an incidence lower than 0.006.

As reflected in the results of this study, the methodology which was developed was successful in distinguishing among fetal and maternal genetic hypotheses in most instances. A selection procedure based on choosing the model yielding the lowest $-LL$ resulted in an error rate of 12% and 6% under both fetal genetic simulations, respectively. The maternal genetic model was always selected when it was the true mode of inheritance. The best explanation for the success of the twin kinship design in selecting the correct transmission model lies in the genetic relationships contained in the pedigree structure. Under fetal genetic inheritance, the design utilizes information provided by full sibs, half-sibs and cousins, with corresponding genetic correlations of 0.500, 0.250, and 0.125. Information pertinent to maternal genetic inheritance is provided for individuals sharing genetically identical mothers, those whose mothers are fully siblings and individuals whose mothers are genetically unrelated. These relationships translate into genetic correlations of 1.0, 0.5, and 0.0, respectively. The wide range of genetic correlations under both fetal and maternal genetic transmissions probably accounts for the increased efficiency associated with the twin kinship design. Although the offspring of MZ twins are genetically the same as "true" half-siblings, the latter family structure alone would be not sufficient in an analysis of this sort since the methodology presented here relies on the offspring of both MZ and DZ twins.

The fact that minimizations under the fetal genetic model did not absolutely locate the global minimum, which resulted in final parameter estimates that were somewhat dependent on the initial values used, would tend to indicate that the likelihood surface was somewhat flat because of insufficient information to adequately estimate all of the parameters. However, the minimization under the maternal genetic model was capable of locating a single minimum, presumably the global minimum. A possible explanation for this is that more information is derived from twin kinships generated under a maternal genetic model than under a model of fetal genetic etiology. Another disappointing feature of the procedures developed here was the inability to accurately estimate parameter values, although many of the estimates were fairly close to their real values.

The results obtained from this study suggest that the models constructed are capable of effectively discriminating among the etiologies evaluated. Although their use in estimating fetal or maternal genetic parameters is limited, this method is superior to those previously used to detect inherited maternal effects on discrete traits [9,13]. Also, other methods of pedigree analysis used to investigate discrete traits not only have had difficulty in estimating parameters, but quite often could not distinguish among competing hypotheses [3,8]. Although the set of competing hypotheses evaluated in this study were different from those examined in other studies, the high power of the methods presented here represents not just a reflection of that difference, but more likely results from the use of data that are more informative and relevant to the hypotheses being tested. Indeed, the level of power in tests of classic genetic hypotheses obtained from using data on the offspring of twins was greater than that associated with other sampling units [5,10]. Although the sample sizes used in this study were greater than for previous simulation studies, there were actually fewer informative kinships.

The results from simulations with an unbalanced design did not differ much from those obtained using balanced data. It has been suggested that for some statistical tests, such as an analysis of variance or a χ^2 for a 2×2 table, a balanced design yields the highest power [4,6,12]. This was not the case here, indicating that the methodology presented can discriminate among different modes of inheritance even when twin parents are not equally distributed for sex by zygosity categories.

The present study contains some caveats that bear mention. Genetic heterogeneity may result in incorrect parameter estimates, as well as impair the resolution of conflicting hypotheses. The variable expression of birth defects, as well as their imprecisely defined transmission may indicate the existence of many loci, each capable of causing a type of the disease. There could also be selection occurring against parents who had the anomaly, such that their contribution to the next generation is reduced. In that case, the genotype frequencies would not be in Hardy-Weinberg equilibrium and one of the assumptions of the model would be violated. Another type of selection that might occur arises from the truncation of family size after one or two affected individuals. Also, since the data used is limited to live births, the gene frequency for the anomaly and the penetrance functions might be underestimated if there is prenatal selection against affected fetuses. Finally, the possibility exists of a correlated environmental effect mimicking genetic transmission.

Data on birth defects do not have many of the problems that may be encountered in the analysis of data on quantitative traits. Assortative mating is likely to be small or nonexistent for certain malformations. Since all individuals are measured at birth, there is no temporal variation of gene expression.

There are many possible avenues of research emanating from this study. It would be desirable to test a mixed fetal and maternal genetic model. In addition, the types of inheritance could be increased to include polygenic inheritance of both fetal and maternal genetic effects. Common family environment might also be included, but it would be comprised of nongenetic elements of the fetal environment, such as maternal nutrition, alcohol intake and smoking. The flexibility of the logistic risk function allows for the addition of environmental variables and/or marker loci. Another elaboration of this study would be to extend the likelihood calculations to include the twins and spouses themselves, as well as their families, if disease status information is available on these individuals. This would either restrict the traits studied to mild afflictions or require the inclusion of a selection factor.

CONCLUSION

The use of MZ and DZ twin kinships in pedigree analysis was examined to see whether this sampling scheme provided an efficient mechanism for detecting inherited maternal effects on birth defects. From the results obtained here, it appears that the twin kinship design was quite powerful in correctly distinguishing among single locus fetal and maternal genetic effects. With the modelling of complex risk functions, and the inclusion of polygenic inheritance, twin kinship may represent a valuable tool in quantitative genetic studies of congenital anomalies.

REFERENCES

1. Bjerkreim I (1976): Congenital dislocation of the hip in Norway: A clinical-epidemiological study. Oslo City Hosp 26:76-90.
2. Corey LA, Winter R, Eaves LJ, Golden W, Nance WE (1980): In Melnik M, Bixler D, Shields E (eds): Etiology of Cleft Lip and Palate. Alan R. Liss, Inc., New York.
3. Demenais F, Elston RC, Bonaiti C, Briard ML, Kaplan EB, Namboordiri KK (1981): Segregation of congenital glaucoma: Approach by two different models. Human Genet 33:300-306.
4. Fleiss JL (1981): Statistical Methods for Rates and Proportions. John Wiley and Sons, New York.
5. Go RC, Elston RC, Kaplan EB (1978): Efficiency and robustness of pedigree segregation data. Am J Hum Genet 30:28-37.
6. Hogg RV, Craig AT (1968): Introduction to Mathematical Statistics. MacMillian Co., New York.
7. Kleinbaum DG, Kupper LL, Chambless LE (1982): Logistic regression analysis of epidemiologic data: Theory and practice. Communications in Statistical Theory and Methodology II: 485-547.
8. Lalouel JM, Morton NE, Jackson J (1979): Neural tube malformations: Segregation analysis and calculation of recurrence risks. J Med Genet 16:8-13.
9. Lalouel JM, Morton NE, MacLean CJ, Jackson J (1977): Recurrence risks in complex inheritance with special regard to pyloric stenosis. J Med Genet 14:408-414.
10. MacLean CJ, Morton NE, Lew R (1975): Analysis of family resemblance IV: Operational characteristics of segregation analysis. Am J Hum Genet 27:365-384.
11. Nance WE, Corey LA: Genetic models for the analysis of data from the families of identical twins. Genetics, 83:811-826.
12. Neter J, Wasserman W (1974): Applied Linear Statistical Models. Richard D. Irwin, Inc., Homewood, IL.
13. Winter RM, Golden WE, Nance WE, Eaves LJ (1978): A halfsib model for the analysis of qualitative traits. Am J Hum Genet 30:129A.

Correspondence: Dr. Andrew A. Kramer, Department of Social and Preventive Medicine, 2211 Main Street, Buffalo, NY 14214, USA.